Mem. Fac. Eng., Osaka City Univ., Vol. 39, pp. 95−98 (1998)

An Appropriate Number of Hidden Units for Three Layered Neural Networks to Parity Discrimination

Jiro OKAMOTO [*], Shigeyoshi NAKAJIMA [†], Masato Asai [‡] and Shoichi HOSOKAWA [§]
(Received September 30,1998)

Synopsis: When we are going to design a neural network(NN) to discriminate input parity, it is always a difficult problem how many numbers of hidden units to be chosen. There are scarcely certain methods to decide them. They are nowadays selected by ones experience and intuition.

In this paper, we describe the invested result about the suitable number of hidden units when the parity of N bit signal is discriminated. The reason why we choose the parity discrimination problem as our research subject is that this problem is a fundamental one and is one of the most difficult problem for binary signals.

We conclude that, in the case of the forward type three layered NN's, there exists the most appropriate number of hidden units for discriminating binary signals.
Keywords: *neural network, parity discrimination, the number of hidden units, binary input, binary signal, computational efficiency, learning, hidden layer*

1 Introduction

In this paper we take a problem of how many number of hidden units is appropriate concerning the parity discrimination. Here we only treat the forward three layered all connected type neural networks (NN's).

About the problems concerning the minimum number of hidden units, several papers have been published including our previous papers [1]∼ [7]. But there are few papers describing the appropriate number of hidden units concerning parity discrimination [8].

The reason why we take the parity discrimination problem as our subject is that (1)the parity discrimination problem is a fundamental one for binary input discrimination, and (2)the minimum number of hidden units has been proved to be $[N/2]+1$ for the case of N bits parity discrimination [4] [6] [7]. Where $[*]$ indicates the Gaussian symbol.

One of the main result in this paper lies in the clarification that there exists an appropriate number of hidden units by which an efficient learning is done and that the number is experimentally shown to be proportional to the input bit numbers, when the input number exceeds four.

When we could find the rules between the minimum number of hidden units and the complexity of the problem concerning parity discrimination, we may extend the way of the thinking to the other problems and will be able to conjecture the appropriate number of hidden units for them.

We think that the number of the minimum hidden units is closely connected to the inner expression of a given problem and it is concerned to the complexity of the problem, so we think, to know the complexity of the problem is to know the inner expression as well as to know the minimum number of hidden units and then the appropriate hidden units.

* Associate Professor, Department of Information and Computer Engineering
† Research Associate, Department of Information and Computer Engineering
‡ Master Course Student, Department of Information and Computer Engineering
§ Professor, Department of Information and Computer Engineering

2 Parity Discrimination

There are 2^N different combinations of signals for N bits. To discriminate these signals by parity is to classify them into two categories in which adjacent vertexes to each other are classified into two different categories respectively, that is, plus and minus parity, considering the signals as the vertexes in a N dimensional hyper cube. This classification is regarded as one of the most difficult classification problem for binary signals.

The minimum number of hidden units for a forward type three layered NN to discriminate N bits parity has already been proved to be $[N/2]+1$ as mentioned above. Where $[*]$ is the Gaussian symbol and takes the integer part of the real number $*$.

Though this number is theoretically proved to be the possible minimum number of hidden units to discriminate input parity, this number is actually an impossible number to discriminate input parity when the input number N exceeds seven or more. Accordingly to know how many times of hidden units compared to the theoretical minimum number is appropriate, is thought to be very useful and necessary, even for other discrimination problems. Because it makes possible to estimate the appropriate number of hidden units by knowing the number of the inner expression of the given problem which may be the minimum number of hidden units.

In order to obtain the appropriate number of hidden units for parity discrimination, we will define the next evaluation quantities as;

$$
\begin{aligned}
&\textit{The computational efficiency} \\
&= \frac{\textit{The convergent rate to the global minimum}}{\textit{The average value of computing cycles until its convergence}}
\end{aligned}
\tag{1}
$$

This definition is based on the idea that the convergent rate is gradually saturated when the number of hidden units is increased, but the computational time to the convergence is much more increased than the increasing rate of hidden units, so the efficiency is highest at the point when this ratio, that is, this computational efficiency, is maximum. Also this definition is an indicator of the convergent efficiency to the global minimum.

Accordingly this value indicates the magnitude of the convergent speed to the global minimum, that is, the efficiency of NN, as the larger the value the shorter the convergent time to the global minimum. In the sequel, we name the number of hidden units when this computational efficiency is maximum as the *appropriate number* of hidden units .

3 Experiments and their results

Fig. 1 shows the relation between the computational efficiency on the vertical axis and the number of hidden units on the horizontal axis ranging from 2 to 30. The variable parameters for NN's in this experiment are (1) 0.2 for the learning rate, (2) 0.9 for momentum ,(3) 20,000 trials for learning cycles and (4) 1.0 for temperature. All experiments in this section are done by the same conditions as this. The computational efficiencies are obtained by the average values of ten trials for each point in the graphs. The computational efficiency is indicated by the normalized values, that is, all values are divided by the individual maximum value for each graph, so that the maximum value is one for each graph.

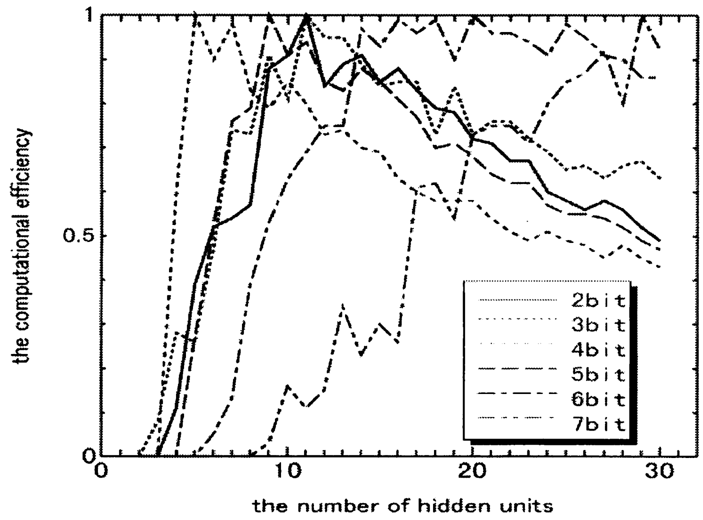

Fig. 1 The relation between the number of hidden units and the computational efficiency when the input bits are changed from 2 to 7.

In the parity discrimination experiments done for two to seven input bits, the numbers of hidden units are changed from two to thirty as described above.

By observing each graph in Fig.1, we can easily see that the computational efficiency becomes larger as the number of hidden units becomes larger, then saturates at a point near between four and ten times of the input bits and then gradually decreases for all trials.

Accordingly we can easily see that there exists the most appropriate number of hidden units concerning the computational efficiency for all trials.

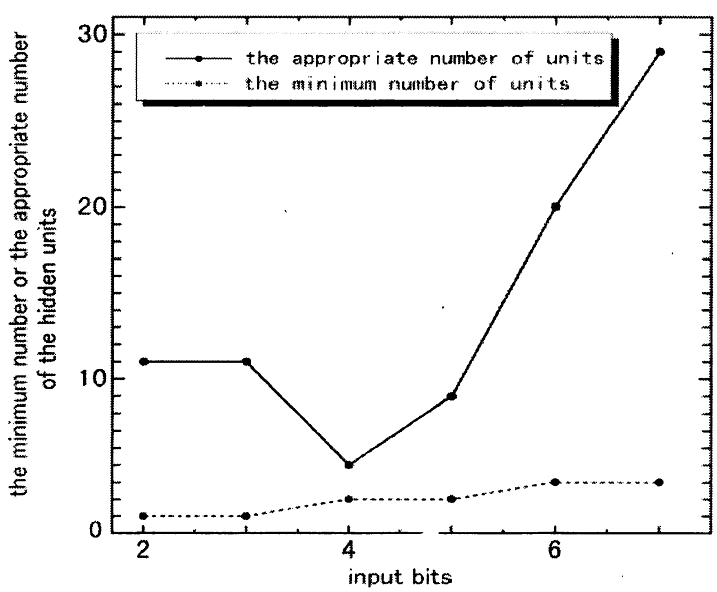

Fig.2 The relation between the number of input bits and the number of hidden units at the points of the maximum computational efficiency and the possible minimum number of hidden units

Fig.2 shows the relation between inputs bits on the horizontal axis and the theoretically minimum number and the appropriate number of hidden units on the vertical axis.

Judging from above two graphs, we could conclude that the appropriate number of the hidden units exists and it is proportional to the input numbers over a point exceed four. These relations may be expressed as the next equation as,

$$y = 10x - 40 \quad (x \geq 4), \tag{2}$$

where y and x indicate appropriate number of hidden units and input bits respectively. But we must proceed more experiments to get more accurate one by increasing input bits.

4 Conclusion

For the forward type three layered NN, we could show that there exists the most efficient number of hidden units which is proportional to the number of input bits when it exceed four and we name it as the appropriate number of hidden units, at least concerning the parity discrimination problem. But the task to get more accurate values of the proportional coefficient is left as our future problem.

One of the other future subjects is to obtain the most efficient number of hidden units for general problems, that is to say, to clarify what times of the minimum number of hidden units is suitable, which may be connected to the inner expression of NN and may be obtainable by the principal value analysis to the given problem.

Throughout these experiments we think that the parity discrimination may play a role like the benchmark test for the construction of NN as for von Neumann type computer.

Acknowledgment

We should like to express our thanks to master course student Mr.M. Hashimoto for his pioneering development of a program for this experiment.

References

[1] D.E. Rumelhart and J.L. McCelland(eds); Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Chap.8, MIT Press, 334/335(1986)

[2] E. Milotti; Parity and the Shapes of Neuron Transfer Functions in Feed forward Neural Networks, Int. J. Neural Systems (Singapore), **1**, 3, 247/248(1990)

[3] D.G. Stork and J.D. Allen; How to solve the N-bit Parity Problem with Two Hidden Units, Neural Network, **5**, 6, 923/926(1992)

[4] J. M. Minor; Parity with Two Layered Feed Forward Nets, Neural Networks, **6**, 5, 705/707(1993)

[5] Okamoto J. and Kawamura Y.; On the Hidden Unit Numbers Necessary for 3 Bit Parity Discrimination, Trans. IEICE, **J76-DII**, 8, 2146/2147(1993)

[6] J. Okamoto and S. Hosokawa; On the minimum Hidden Unit Numbers to Discriminate Input Parity. Trans. IEICE, **J77-DII**,9,1956/1959(1994)

[7] Rudy Setiono; On the solution of the Parity Problem by a Single Hidden Layer Feedforward Neural Network, Neurocomputing, vol.16, 225/235(1997)

[8] J. Okamoto, S. Nakjima and S. Hosokawa; Learning Efficiency and Temperature Coefficient of Forward Type Three Layer Neural Networks, Memoirs of the Faculty of Engineering Osaka City University, vol.38, 183/191(1997)